

Difference to Inference: Teaching logical and statistical reasoning through on-line interactivity

THOMAS E. MALLOY

University of Utah, Salt Lake City, Utah

Difference to Inference is an on-line JAVA program that simulates theory testing and falsification through research design and data collection in a game format. The program, based on cognitive and epistemological principles, is designed to support learning of the thinking skills underlying deductive and inductive logic and statistical reasoning. Difference to Inference has database connectivity so that game scores can be counted as part of course grades.

Emphasizing the active nature of information processing, Posner and Osgood (1980) proposed that computers be used to train inquiry in a way that "the teaching environment . . . arise[s] in close relationship to the inquiry activities of the faculty who shape it" (p. 95). Recent examples of educators designing computer-assisted learning contexts in which students actively discover knowledge by thinking like the practitioners of a discipline include examples as diverse as designing research studies (Brown, 1999) and solving ethical dilemmas in medicine through case studies (Martin & Reese, 2000). Washburn (1999), focusing on one important mental skill of scientists, uses an interactive program to teach students, who typically accept published inferences as fact, to distinguish between factual statements and inferences in research reports. The Difference to Inference game engages students in one fundamental activity of scientific inquiry—the use of deductive logic, inductive logic, and statistical reasoning in the falsification of theories.

Game Procedure

Accessing Difference to Inference. The game procedure is laid out in detail in StatCenter at the University of Utah (<http://www.psych.utah.edu/stat/introstats/>). It is free and freely available to instructors and students. A visitor may log in as a guest. Scores earned by one guest overwrite scores earned by other guests. Teachers wanting to keep individual scores for students may contact the author to make an arrangement for doing so.

On the main menu, under "Work and Learn," select "Interactive Learning." Scroll down to "Difference to In-

ference" and click. There will be a choice of "stories" that give meaning to game activities. Different instructors prefer to frame the game with different stories to reflect the content of their courses. For this discussion, "Hurricane Damage" will be used as a frame. Selecting "Hurricane Damage" will show a menu with historical notes about hurricane damage and a "Start the Game" option. The game interface comes with an extensive tutorial and with a link to an on-line lecture integrating game play into statistical theory. Space does not allow for more than an overview of the game here.

Game story. Students are asked to plan a strategic series of two-group experiments whose data will discriminate among five candidate theories. They are shown five maps that describe the pattern of deforestation in five provinces of a tropical country prior to a hurricane. An example of a set of five patterns of deforestation is shown on the far right of Figure 1. The dark gray cells in each of the five maps represent an area of deforestation; the light gray cells represent areas that remain forested. (On a computer screen, the dark gray cells are red, and the light gray cells are green.) The records indicating which pattern of deforestation goes with which province have been lost during the hurricane. The students are to suppose that they are in one of the provinces and have been given a grant to determine which of the five maps applies to that province. It is known that mortality because of mud slides is higher in areas of land that are deforested than in areas protected by rain forest. The students are to use hurricane mortality data to determine which deforestation map applies to the province in which they are located. The game has five levels of difficulty. Difficulty level depends on the effect size of deforestation: Does deforestation produce increases in mortality that are enormous and require no statistical analyses (easiest level), or are they more subtle, perhaps so subtle that even powerful inferential statistics such as *t* might be insufficient and replication is required (hardest level).

Interface. Figure 1 shows the game interface. Five candidate theories are in a column on the right. The family of candidate theories is randomly generated by a schema

This project was supported by a Utah Higher Education Technology Initiative grant. The JAVA programming for the Difference to Inference game was designed and written by Gary Jensen with Matthew Graham. Gary Jensen also designed the Oracle database and its interaction with the JAVA applet. The Malloy (2000) paper cited in the references was based on the same JAVA applet but addressed different theoretical issues. Correspondence regarding this article should be addressed to T. E. Malloy, University of Utah, Department of Psychology, 380 S. 1520 E., Room 502, Salt Lake City, UT 84112-0251 (e-mail: malloy@psych.utah.edu).

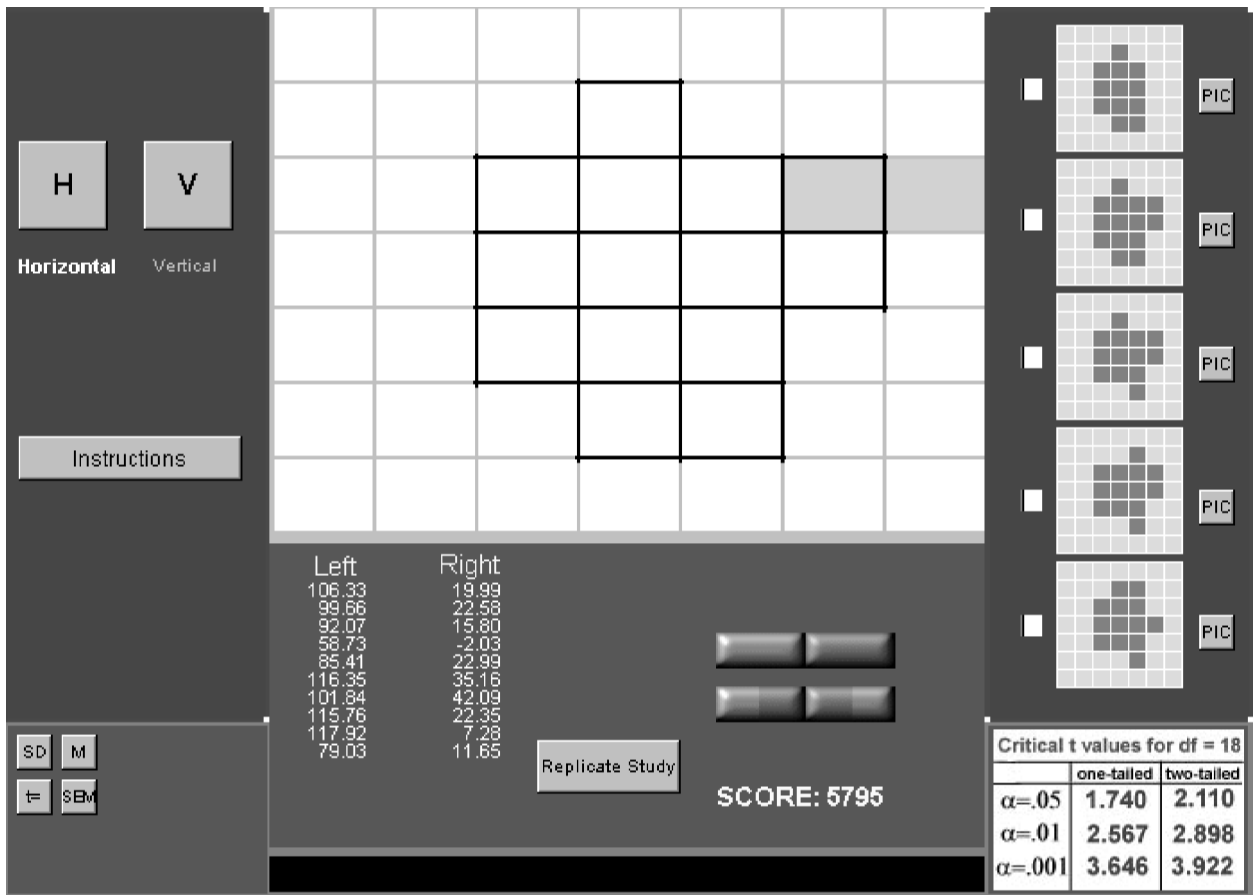


Figure 1. Interactive Graphical User Interface for Difference to Inference game. Students must discover which of the five patterns in a column on the right is hidden behind the white grid in the center. The data (shown at the bottom) have been collected from selected cells (shown in gray).

plus exception rule for each student each time the game is played. The white grid in the center of the interface is a workspace that allows students to use horizontal or vertical selection tools to choose to see mortality rates for any two adjacent cells.

Buttons next to each candidate theory allow the outline of the theory to be projected onto the white workspace. Figure 1 shows the outline of the second-from-the-top theory projected onto the workspace. Allowing a theory to be outlined on the workspace keeps the task from primarily requiring visual-spatial skills and lets players focus on logic. The example in Figure 1 shows a case in which a player has used a tool to select two cells (highlighted in gray) on the far right of the work area. Notice that, in the column of theories, the top and bottom theories show both these two cells to be forested; therefore, those two theories predict that both cells would have low mortality rates (no difference). In contrast the middle three theories in the column show that, of the two cells selected by the player, the one on the left is deforested, and the one on the right is forested. Therefore, the middle three theories predict that there should be high mortality rates in the left cell and low mortality in the right cell (difference). These different predictions are differences that

make a difference: Depending on the data, either three of the theories or two of the theories will be falsified.

In short, players can query mortality rates from any two adjacent cells. Data resulting from the query appear below the white workspace. The data is generated pseudo-randomly from Gaussian distributions whose parameters are set when difficulty level (deforestation effect size) is set; thus, each data query results in a unique data set. Buttons allow the students to ask for statistical analyses, such as group means, standard deviations, standard errors (*SEMs*), and a *t* test.

Scoring. The students start the game with a small grant. Collecting and replicating data costs money, as do statistical analyses performed on collected data. After conducting a series of two-group studies, the students select one theory (map) as the best description of the pattern of deforestation on the basis of the hurricane mortality data. The program is structured so that only one of the five candidate theories is consistent with the statistical equations that generate the data sets. The other four theories will be inconsistent with the data—at least, they will be inconsistent with the data if the right research questions are asked. The students' choice of theory is recorded on an Oracle database and affects their course

grades. Submission of a theory that is inconsistent with the data costs a great deal of grant funds. After submitting a theoretical conclusion, the students can play the game again with five new (randomly generated) theories. This process continues until the students earn large enough grants to get full credit for that particular level of the game. They then begin on the next level, in which treatment effects are more subtle and statistical reasoning more important.

Game Design: Difference to Inference

Deducing differences that make a theoretical difference. Students must deduce a specific research question on the basis of a critical examination of the five deforestation patterns. If they are to eliminate theories on the basis of mortality data, first they must seek the edges of the deforestation patterns where data collected in a two-group study will show differences in mortality. But not all such differences are theoretically important. Comparative examination of the edges of the five theories shows that all five make different predictions for many comparisons. A research study that finds such a difference cannot discriminate among theories. So students must decide where to set up a two-group comparison whose data might eliminate some theories, but not others. They must not just find differences between their groups, but they must find those “differences that make a difference” in eliminating theories (Bateson, 1972, p. 452). This can be done by critically examining the five maps on the right and deducing where the theories make different predictions. The two selected cells (shown in gray) in Figure 1 target a difference in mortality that will make a difference in eliminating theories.

Deduction and falsification. The data obtained from research studies may or may not be inconsistent with each theory’s prediction. If a theory’s prediction is shown not to be true, the theory is falsified through deductive logic. In the Difference to Inference game, the data from a well-designed study can be used to falsify one or more of the theories.

The fundamental inductive pattern. In contrast, data that are consistent with a theory’s prediction do not prove the theory. Yet, the empirical scientific method can be persuasive when theories accurately predict data, especially across a series of studies. Polya (1968) develops the important distinction between deductive falsification and inductive plausibility. When the prediction of a theory is found to be false, the theory is proven to be false. But when the prediction of a theory is verified, the theory is not proven to be true; rather, it becomes more plausible.

Polya describes *inductive patterns* by which theories, although not proven, become more and more plausible. The fundamental inductive pattern is the verification of a consequence of a theory. When a prediction of a theory is found to be true, the theory becomes more plausible. Other inductive patterns that lead to increased plausibility include successive verifications of a theory’s predictions and the elimination of rival theories (Polya, 1968,

p. 26). The Difference to Inference game gives students the opportunity to learn about induction and plausibility. They gain repeated experience with a theory that, although not proven, becomes more plausible, because its predictions are always verified, whereas rival theories are eliminated. Such experience provides a basis for discussions of the distinction between deductive falsification and inductive plausibility in class.

Statistical reasoning. Up to this point, the game has been discussed as if the data were clear cut: Either the data from the two targeted cells showed a difference in mortality or not, and this difference (or lack of it) either was consistent with a particular theory’s prediction or it was not. The emphasis has been on the use of deductive and inductive logic. But at higher levels of game difficulty, the treatment effects are small, and the data obtained for the two targeted cells can show a great deal of overlap. It may not be at all obvious whether or not there is a difference in mortality in the two cells. At these levels of difficulty, the game requires statistical reasoning, including the use of null-hypothesis-testing logic using procedures like a *t* test. This allows students, after they have learned about the basic logic of theory falsification, to learn to integrate statistical reasoning into the research process.

As one component of StatCenter, the Difference to Inference game is supported by several HTML lectures on distribution theory, sampling distributions, hypothesis testing, *t* tests, and other common tests. So students have a rich context for learning statistical reasoning when they play the most difficult levels of the game. Consequently, at high-difficulty levels, the game gives students practice integrating statistical reasoning with deductive and inductive reasoning to make theoretical inferences. Both Cook and Campbell (1979) and Polya (1968) refer to this step in the scientific hypothesis as eliminating the plausible competing hypothesis of chance. Higher difficulty levels of the game facilitate formal class discussions of how chance alone may account for the differences in two sets of data and the importance of eliminating chance as a rival theory.

Final inference. At the end of a series of studies, students must choose a theory. The game is structured, somewhat artificially, so that only one of the five candidate theories will remain consistent with the simulated data from all the possible comparisons of the mortality rates in two adjacent cells. The other four theories will be eliminated by the results of at least one possible research study. To succeed in the game, the students must critically design and conduct a series of two-group research projects whose data eliminate four of the theories. They are to choose the theory that remains consistent with all the known data. This inference is very different from inferring that a theory has been proven. Often, students enter a methods course with preconceptions inclining them to want to prove theories true. The Difference to Inference game addresses this teaching need by giving students repeated experience in eliminating theories that are inconsistent with data and preferring theories that gain plausibility through their consistency with known data.

Discussion

The program structure of Difference to Inference is based on principles from logical and plausible reasoning (Polya, 1968), statistical reasoning (Cook & Campbell, 1979; Polya, 1968), and epistemological frameworks about the nature of information and pattern (Atneave, 1954; Bateson, 1972). Using these principles, the game simulates empirical scientific theory. Students design studies whose data eliminate one or more theories until only one theory remains that is consistent with the known data.

The use of visual patterns as theories has several advantages. Atneave (1954, p. 187) noted that there is a similarity between the "abstraction of simple homogeneities from a visual field" and the "induction of a highly general scientific law from a mass of experimental data." Using this similarity, a computer can generate families of competing but overlapping visual theories in ways that make deductions of research designs challenging. This allows students to gain repeated practice with the processes of theory elimination across many games without burdening the instructor with the need to construct endless examples of competing theories. The simple visual nature of the patterns also allows for the game structure to teach the targeted principles (such as finding differences that make a difference), using unambiguous materials.

Malloy (2000) has sketched a general framework, based on procedural or nondeclarative memory (Squire, Knowlton, & Musen, 1993), that hypothesizes that repetitive experience with computer games builds procedural knowledge. The nature of that knowledge depends on the logic of the computer game's program. Even setting aside the issue of violent content, most commercial games have trivial logical structures, and the resulting procedural learning will be equally trivial. But educators can build games whose logic is based on principles valued by various disciplines, such as those underlying the process of scientific inquiry. Difference of Inference was designed to require repetitive experiences, which is consistent with the logic of scientific inference and might therefore lead to procedural knowledge important to scientific thinking.

The advantage of visual patterns as theories leads directly to a primary limitation of the Difference to Inference game. Whereas most theories in the social sciences are verbal and complex, sometimes even ambiguous, the game uses simple, clear-cut visual patterns as candidate theories. There are, obviously, questions of relevance and generalizability. This limitation can be addressed in the context of a course in which the instructor provides the bridge to appropriate content to aid in the generalization of thinking skills learned while playing the game. It is important to integrate procedural and declarative learning. Although Difference to Inference explicitly emphasizes the learning of the research process, it is best used in a context that provides the substantial declarative knowledge support of a typical methods or statistics course. One of the limitations of declarative knowledge is that it is difficult to learn about processes declaratively without

experiencing the process. Reading about how to ski typically provides more insights after a few runs down the mountain than it does prior to such experience. In the same way Difference to Inference and other simulation software that give practice in scientific thinking can give students many runs through the structure of the research experience to provide a foundation for subsequent declarative knowledge in lectures and text. For example, the author, for one, finds it much easier to lecture about the difference between proving theories and falsifying theories after, rather than before, students have had experience with the falsification process in Difference to Inference. In short, Difference to Inference is not meant to stand alone but to be used in a supportive declarative context. In that way, students' repeated experience with the logical and statistical reasoning processes involved in theory falsification provides a concrete basis for declarative knowledge statements about the philosophy of science. The experience provided by Difference to Inference can be generalized to verbal research questions based on verbal theories by using StatCenter's Virtual Lab applet (Malloy & Jensen, 2000).

Difference to Inference is one example of a general proposal (Malloy, 2000; Posner & Osgood, 1980): The logical structure of computer programs can be designed to correspond to the structure of culturally valued activities (e.g., empirical scientific research) so as to aid students in developing internal thought structures for accomplishing those valued activities.

REFERENCES

- ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychological Review*, **61**, 183-193.
- BATESON, G. (1972). *Steps to an ecology of mind*. New York: Ballantine.
- BROWN, M. F. (1999). Wildcat World: Simulation programs for teaching basic concepts in psychological science. *Behavior Research Methods, Instruments, & Computers*, **31**, 14-18.
- COOK, T. D., & CAMPBELL, D. T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.
- MALLOY, T. E. (2000). Teaching deductive, inductive, and inferential logic through interactive online computer simulation. *Journal of Information Technology in Medicine* [On-line], **3**. Available: <http://www.J-ITM>
- MALLOY, T. E., & JENSEN, G. C. (2000, November). *Utah virtual lab: JAVA interactivity for teaching science and statistics online*. Paper presented at the meeting of the Society for Computers in Psychology, New Orleans.
- MARTIN, R. M., & REESE, A. C. (2000). Computer assisted instruction as a component of a comprehensive curriculum in medical ethics. *Journal of Information Technology in Medicine* [On-line], **3**. Available: <http://www.J-ITM.com>
- POLYA, G. (1968). *Patterns of plausible inference*. Princeton, NJ: Princeton University Press.
- POSNER, M. I., & OSGOOD, G. W. (1980). Computers in the training of inquiry. *Behavior Research Methods, Instruments, & Computers*, **12**, 87-95.
- SQUIRE, L. R., KNOWLTON, B., & MUSEN, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, **44**, 453-495.
- WASHBURN, D. A. (1999). Distinguishing interpretation from fact (DIFF): A computerized drill for methodology courses. *Behavior Research Methods, Instruments, & Computers*, **31**, 3-6.